La Vision Artificielle

Par Radu HORAUD

LIFIA (Laboratoire d'Informatique Fondamentale et d'Intelligence Artificielle) - INPG, 46, avenue Félix-Viallet, 38031 GRENOBLE CEDEX

1. Qu'est-ce que la vision ?

L'espace qui nous entoure a une structure tri-dimensionnelle (3D). Lorsque l'on demande à une personne de décrire ce qu'elle *voit*, elle n'éprouve aucune difficulté à nommer les objets qui l'entourent : téléphone, table, livre... Et pourtant l'information qui est réellement disponible sur la rétine de ses yeux n'est ni plus ni moins, une collection de points (environ un million !). En chaque point ou *pixel* (picture element) il y a tout simplement une information qui donne une indication quant à la quantité de lumière et la couleur qui proviennent de l'espace environnant et qui ont été projetées à cet endroit de la rétine. Le téléphone, la table ou le livre *n'existent pas* sur la rétine. Guidé à la fois par l'information codée dans l'image (ou la rétine) et par ses propres connaissances, le processus visuel *construit* des percepts. Le téléphone ou le livre sont les réponses finales, résultant d'un processus *d'interprétation* qui fait partie intégrante du système de vision. De plus, il n'y a pas de correspondance terme à terme entre l'information sensorielle (la lumière et la couleur) et la réponse finale (des objets 3D). Le système de vision doit *fournir* les connaissances nécessaires afin de permettre une interprétation non ambiguë.

2. Comprendre la vision

Il n'est pas suffisant de constater qu'un problème est complexe. Encore faut-il essayer de le comprendre dans ses moindres détails et proposer une solution.

La vision a suscité l'intérêt de nombreux scientifiques et philosophes depuis déjà très longtemps. Parmi ceux-ci, les neurobiologistes mènent des recherches théoriques et expérimentales afin d'essayer de comprendre l'anatomie et le fonctionnement du cerveau dans son ensemble. Ils ont découvert une structure très complexe qui est loin de leur avoir révélé tous ses secrets. La tâche des neurobiologistes semble être à la fois grandiose et illusoire. Grandiose, parce que le cerveau est une des plus complexes *inventions* de la nature. Il reste et restera pour longtemps le bastion encore inconnu que les sciences humaines se proposent de conquérir. Illusoire, car on ne connaît pas ses limites. Ces limites ne sont-elles pas repoussées à chaque découverte ? David Hubel [1] a merveilleusement bien exprimé ce paradoxe : *Le cerveau peut-il comprendre le cerveau ?*

Avec la naissance de machines de calcul de plus en plus sophistiquées, un certain nombre de scientifiques se sont attaqués au problème de la vision d'un point de vue quantitatif : est-il possible de construire un *modèle computationnel* pour la perception visuelle ? Attention : il ne s'agit pas de fournir une explication de comment marche la vision biologique mais de **créer un modèle** qui vu de l'extérieur possède des propriétés semblables.

Ce modèle *artificiel* peut-il être d'une utilité quelconque quant à la vision biologique ? Peut-il constituer la base d'une nouvelle technologie – *des machines qui voient ?*

3. Une approche scientifique

Il est certainement trop tôt pour répondre à ces questions et pour tirer des conclusions. Malgré les efforts non négligeables, il y a très peu de résultats convaincants. Nous pensons que deux démarches doivent être suivies simultanément :

Essayer d'élaborer une théorie de la **vision artificielle** qui doit nous guider à long terme.

Tenter de résoudre des problèmes spécifiques dans le cadre de cette théorie : de tels résultats partiels permettraient de confirmer ou au contraire de mettre en cause certains aspects de la théorie.

Nous préférons utiliser la nomination « vision artificielle » à la place de « vision par ordinateur ». Ce dernier terme implique l'utilisation d'une machine pourvue d'une architecture conventionnelle du type von Neuman. Restreindre l'implémentation de la théorie à ce type de machines nous parait un peu limitatif. D'autre part le mot *artificiel* illustre bien qu'il s'agit d'un système synthétisé par l'homme qui pourrait par certains de ses aspects imiter les apparences d'un système naturel [2].

4. Une théorie de la vision

L'élaboration d'une théorie scientifique demande trois étapes :

1. Énoncer la théorie. Spécifier et élaborer les concepts de base. Ces concepts doivent exprimer le cadre formel qui est à la base de la théorie.
2. Exprimer ces concepts sous forme mathématique.
3. Réaliser un ensemble expérimental qui permette de vérifier la théorie.

Voici comment la vision artificielle peut s'énoncer brièvement dans les termes de ce paradigme. La vision est un processus de traitement de l'information. Elle utilise des stratégies bien définies afin d'atteindre ses buts. L'entrée d'un système de vision est constituée par une séquence d'images. Le système lui-même apporte un certain nombre de connaissances qui interviennent à tous les niveaux. La sortie est une description de l'entrée en termes d'objets et de relations entre ces objets.

Deux types de stratégies sont mises en jeu : ascendantes et descendantes. Les stratégies ascendantes tentent de construire à partir de l'information sensorielle une représentation la plus abstraite possible (par exemple, un ensemble de primitives géométriques 3D). Les stratégies descendantes déduisent à partir de l'ensemble d'objets connus par le système une description compatible avec les primitives extraites de l'image. Il est alors possible de mettre en correspondance la représentation extraite de l'image avec les descriptions des objets afin de décrire les données sensorielles en termes de ces objets.

Les connaissances mises en jeu sont de trois types : physiques, géométriques et sémantiques. Les lois physiques imposent des contraintes aux signaux lumineux qui partant d'une source, traversent la scène et se projètent sur l'image. La gravitation impose à la scène (et donc à l'image) une structure hétérogène : prépondérance de lignes verticales et horizontales pour ne citer qu'un exemple. La forme des objets (l'ensemble de ses surfaces) et la géométrie de la formation de l'image imposent des contraintes très strictes quant aux structures susceptibles d'être présentes dans

l'image. A un niveau plus élevé, un objet peut être décrit par sa fonction dans le contexte d'un raisonnement symbolique. Cette fonction n'est pas directement mesurable dans l'image. On devrait pouvoir dériver des contraintes sur la forme et l'emplacement d'un objet à partir de sa fonction. Par exemple, le mot *chaise* désigne une classe d'objets réels (un objet réel est un objet qui occupe une place dans l'espace physique). Cependant il y a une grande variété de chaises quant à la forme et à la couleur pour ne citer que deux propriétés. Quelles sont les propriétés communes à toutes les chaises, **mesurables** dans l'image ?

L'étape suivante consiste à exprimer ces stratégies et connaissances dans le cadre d'un formalisme mathématique et à construire les algorithmes correspondants. Les performances de ces algorithmes doivent correspondre aux qualités exigées d'un tel système : *la reconnaissance visuelle doit être fiable et rapide.*

5. Le paradigme de David Marr

Vers la fin des années 70, David Marr [3] a proposé un modèle computationnel pour le traitement et la représentation de l'information visuelle. Voici quels sont les principaux traits de ce paradigme :

- A partir d'une ou de plusieurs images un processus d'extraction de caractéristiques produit une description en termes d'attributs bi-dimensionnels. Ce niveau de représentation est appelé *première ébauche* (primal sketch).
- La première ébauche constitue l'entrée d'un certain nombre de processus indépendants qui calculent des propriétés tri-dimensionnelles locales relatives à la scène. Il s'agit d'une représentation centrée sur l'observateur, appelée *ébauche 2.5D*. Ces processus opèrent sur une séquence d'images (analyse du mouvement), sur une paire d'images (stéréoscopie) ou sur une seule image. Dans ce dernier cas il s'agit de processus d'inférence qui utilisent des connaissances géométriques (analyse des contours), géométriques et statistiques (analyse des textures), photométriques (analyse des ombrages) ou colorimétriques (analyse des reflets).
- L'ébauche 2.5D est mise en correspondance avec des connaissances 3D afin de construire une description de la scène en termes d'objets et de relations entre les objets. Il s'agit maintenant d'une représentation *centrée sur la scène* (la description ne dépend plus de la position de l'observateur).

6. Un colloque et un numéro spécial

Le modèle proposé par D. Marr est intéressant pour au moins deux raisons. Tout d'abord parce qu'il exprime les problèmes posés par la Vision dans les termes d'une science. Ensuite, parce qu'il s'efforce de tenir compte d'un grand nombre de travaux et des résultats théoriques et expérimentaux qui leur sont associés.

Nous avons regroupé dans ce numéro spécial quelques articles qui sont autant de contributions à l'élaboration d'un modèle computationnel le plus complet possible. Cet ensemble d'articles est une sélection opérée parmi les travaux présentés au colloque scientifique « Vision par Ordinateur » qui a eu lieu à Cargèse (Corse) du 22 au 26 septembre 1986. Une trentaine de chercheurs représentant neuf laboratoires y étaient présents. Le colloque a eu lieu dans un magnifique lieu de rencontres scientifiques : l'Institut d'Études Scientifiques (IES) de Cargèse. Voici un bref sommaire des travaux figurant dans ce numéro spécial.

Patrick Rives et Bernard Espiau présentent une technique permettant d'estimer des primitives géométriques tri-dimensionnelles à partir d'une séquence d'images

obtenues à l'aide d'une caméra mobile. L'idée de base du schéma proposé est d'utiliser des mesures instantanées de paramètres bi-dimensionnels au fur et à mesure que le capteur se déplace dans une scène supposée statique. La trajectoire de déplacement est calculée en fonction de la trajectoire *nominale* correspondant à la tâche principale et de la trajectoire entraînant la meilleure estimation de certaines primitives supposées présentes dans la scène.

Patrick Bouthémy s'attaque au problème de l'analyse du mouvement dans une séquence d'images lorsqu'il n'y a pas de restrictions quant au type de mouvement (translation, rotation, homothétie, ...) et lorsque la scène n'est pas rigide : plusieurs types de déplacements peuvent être présents simultanément dans la même image. Dans l'approche retenue, la segmentation et l'estimation sont conjointement abordées, permettant de fournir un ensemble d'indices spatio-temporels de différents niveaux, exploitable ensuite par une tâche d'analyse de scène. L'information la plus riche qui puisse être extraite est un champ structuré des vecteurs vitesse sur toute l'image. Sommairement, cette méthode se décompose en trois étapes : 1. extraction de primitives locales, 2. structuration intermédiaire et 3. estimation du champ des vitesses.

Olivier Monga et Brigitte Wrobel présentent une méthode de segmentation d'une image en *régions*. Une image peut être caractérisée par un certain nombre d'attributs tels que intensité lumineuse, texture, couleur, etc. Deux régions voisines sont définies de telle façon qu'au moins un de ces attributs change d'une région à une autre. La segmentation d'une image est une étape cruciale pour réduire la quantité d'information. Les auteurs définissent tout d'abord formellement la notion intuitive de région. Ils proposent ensuite un algorithme pour décomposer une image en régions selon le critère d'optimalité choisi. Ils étudient également en détail la complexité de l'algorithme ainsi que la structure de données qui lui est associée. Une comparaison des programmes implémentés indépendamment par les deux auteurs et de nombreux exemples complètent cet article.

Dans une série de deux articles, Anick Montanvert et Jean-Marc Chassery présentent une revue de quelques outils géométriques à mettre en œuvre afin de décrire une image segmentée en termes de *structures* bi-dimensionnelles, ainsi que l'utilisation de ces outils dans le cadre d'une application spécifique : la trajectographie sous-marine. Une revue des techniques de base telles que segmentation polygonale de contours, décomposition en éléments convexes d'une région, description d'une forme bi-dimensionnelle, etc., ainsi qu'une bibliographie très complète sont fournies. Ensuite ils présentent l'utilité de la compréhension d'une image en termes de structures géométriques pour le calcul de la trajectoire d'un robot sous-marin dans un environnement accidenté.

Tous les programmes de segmentation d'images manipulent un nombre de paramètres assez élevé. Catherine Garbay présente une approche « système expert » pour déterminer automatiquement ces paramètres. Elle s'intéresse plus particulièrement à la segmentation d'images cytologiques. Il s'agit notamment de modéliser les connaissances utiles au processus de segmentation. Deux types de connaissances sont utilisés : *spécifiques* à l'application envisagée et qui sont donc associés au contenu sémantique des images à traiter et *non spécifiques,* intrinsèques au caractère bas-niveau de la segmentation.

Enfin, José Luis Gordillo s'attaque à l'analyse des images couleur. L'attribut couleur complète et rend plus riche la description d'une image monoculaire mais en même temps les traitements deviennent plus complexes. Un système conversationnel d'analyse d'images couleur est présenté. Cette analyse fournit la meilleure composante couleur devant être utilisée pour l'extraction de contours et permet de caractériser l'image en termes de régions de couleur homogène. On illustre ensuite l'utilisation de la couleur dans le cadre d'une tâche spécifique : identification de fils électriques dans une paire d'images stéréoscopiques.

Avant de vous laisser découvrir vous-même ce numéro spécial, je voudrais remercier vivement tous ceux qui ont contribué au succès de cette entreprise : Marie-France Hanseler de l'IES pour avoir organisé le colloque de Cargèse, tous les participants au colloque pour leur enthousiasme grâce auquel notre rencontre a pris une tournure scientifique passionante, le comité de rédaction de la revue *TS* pour l'accueil qu'ils ont réservé à notre discipline (une revue spécialisé en Vision, en langue française, n'existant pas actuellement), ainsi qu'à l'ensemble des experts qui ont bien voulu, tout en restant anonymes, apporter leurs observations et critiques indispensables à la réalisation d'un document d'une grande qualité. Au nom de tous les auteurs, mes remerciements vont tout naturellement à Jeanne Malbos : sans son travail assidu, ce numéro spécial n'aurait pu voir le jour.

BIBLIOGRAPHIE

[1] D. Hubel, *The Brain, Scientific American,* 241, n° 3, septembre 1979, p. 39-47.
[2] H. Simon, *The Sciences of Artificial,* MIT Press, 1981.
[3] D. Marr, *Vision,* W. H. Freeman and Company, New York, 1982.